

MAGIA²: from miRNA and genes expression data integrative analysis to microRNA–transcription factor mixed regulatory circuits (2012 update)

Andrea Bisognin¹, Gabriele Sales², Alessandro Coppe³, Stefania Bortoluzzi^{1,*} and Chiara Romualdi^{1,*}

¹Department of Biology, University of Padova, Padova 35131, ²Department of Statistical Science, University of Padova, Padova 35121 and ³Department of Comparative Biomedicine and Food Sciences, University of Padova, Padova 35020, Italy

Received March 1, 2012; Revised April 23, 2012; Accepted April 30, 2012

ABSTRACT

MAGIA² (<http://gencomp.bio.unipd.it/magia2>) is an update, extension and evolution of the MAGIA web tool. It is dedicated to the integrated analysis of *in silico* target prediction, microRNA (miRNA) and gene expression data for the reconstruction of post-transcriptional regulatory networks. miRNAs are fundamental post-transcriptional regulators of several key biological and pathological processes. As miRNAs act prevalently through target degradation, their expression profiles are expected to be inversely correlated to those of the target genes. Low specificity of target prediction algorithms makes integration approaches an interesting solution for target prediction refinement. MAGIA² performs this integrative approach supporting different association measures, multiple organisms and almost all target predictions algorithms. Nevertheless, miRNAs activity should be viewed as part of a more complex scenario where regulatory elements and their interactors generate a highly connected network and where gene expression profiles are the result of different levels of regulation. The updated MAGIA² tries to dissect this complexity by reconstructing mixed regulatory circuits involving either miRNA or transcription factor (TF) as regulators. Two types of circuits are identified: (i) a TF that regulates both a miRNA and its target and (ii) a miRNA that regulates both a TF and its target.

INTRODUCTION

MicroRNAs (miRNAs) are fundamental post-transcriptional regulators of several biological processes, whose alterations have been demonstrated in a number of diseases and in almost all human cancers (1–5).

miRNA imperfectly bind the 3'-untranslated region (3'-UTR) of their target mRNAs and may cause translation inhibition and/or mRNA cleavage and degradation (6,7). miRNAs can have multiple targets and a single protein-coding gene can be targeted by multiple miRNAs. In this perspective, the network of post-transcriptional regulatory relationships is expected to have a highly complex connectivity structure.

In silico target identification is based on (i) sequence similarity search, possibly considering target site evolutionary conservation and (ii) thermodynamic stability. However, it is known that the results of target prediction algorithms are characterized by very low specificity (8). This is caused both by the limited comprehension of the molecular basis of miRNA-target pairing and by the context-dependency of post-transcriptional regulation due to the cooperative interactions of different miRNAs. The integration of target predictions with miRNA and gene expression profiles has been recently proposed to improve the detection of functional miRNA-target relationships (9,10). As miRNAs act prevalently through target degradation, expression profiles of miRNA and target genes/transcripts are expected to be inversely correlated.

Although some studies have demonstrated that the regulation of single key targets may largely explain the function of a given miRNA in tumorigenesis or metastasis development (11–13), the mechanisms of RNA regulation

*To whom correspondence should be addressed. Tel: +39 049 827 7401; Fax: +39 049 827 6159; Email: chiara.romualdi@unipd.it
Correspondence may also be addressed to Stefania Bortoluzzi. Tel: +39 049 827 6502; Fax: +39 049 827 6209; Email: stefania.bortoluzzi@unipd.it

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

are emerging to be much more complex than previously expected. miRNA activity should be viewed as part of a complex system where regulatory elements and their interactors generate a highly connected network, which regulates gene expression on multiple levels. Unfortunately, from an experimental point of view, the cooperation of different miRNAs acting on more than one target and the interplay with other regulatory levels are still challenging to be proved.

The gene regulation at the transcriptional level is, in general, quantitatively stronger than that occurring post-transcriptionally. Both levels impact on mRNA/genes expression profiles. Indeed, the miRNAs expression level can be activated or repressed by transcription factors (TFs), whereas mRNAs encoding TFs can be silenced by miRNAs. miRNAs and TF can share targets. Thus, miRNAs and TFs can form feedback or feed-forward loops, cooperating to tune gene expression (14) and playing critical roles in various biological processes. In a seminal review, Inui *et al.* (15) describes the role of miRNAs in pathways as decision maker elements to discriminate real versus too weak or too transient signals. In this context, the association between miRNAs and target expression profiles is generally more complex than linear correlation.

Although many researchers have attempted to understand the mechanism of miRNAs to regulate target genes and their roles in various diseases, the study of miRNA regulation by TFs (TF–miRNA regulation) has been relatively limited (16). Some attempts have been reported in order to computationally reconstruct regulatory circuits concerning miRNA–TFs and their joint targets (17–19). However, the identification of miRNA promoters is a challenging task. As promoter regions could be several thousands of base pairs upstream of the primary-miRNA, a correct way of identifying putative miRNA regulatory regions was not clearly established so far. To this aim, a few databases have been developed to store experimentally validated TF-miRNAs interactions (16,20).

Although several web tools (21–27) were recently developed in order to enhance the functional insights of miRNA functions and to improve miRNA–mRNA interactions identification almost none of them face the complex problem of the reconstruction of the above-mentioned TF–miRNA–target circuits.

Here, we present MAGIA² (miRNA and genes integrated analysis 2012 update, freely available at <http://gencomp.bio.unipd.it/magia2>) an updated and fully redesigned release of MAGIA (21), a web application for the integrative analysis of *in silico* target prediction, miRNA and gene expression data. As in the previous release, MAGIA² refines *in silico* target prediction through miRNA–target expression association measures. MAGIA² extends the analysis, supporting multiple organisms (human, mouse, rat and drosophila), and using a greatly expanded list of target predictions algorithms. It allows the combination of multiple predictions and the selection of user-defined thresholds for each one. The MAGIA² architecture was completely redesigned: both data and analyses results are stored in a user-specific

environment keeping the data private. Furthermore, the updated MAGIA² tries to dissect regulatory complexity reconstructing mixed regulatory circuits involving either human miRNA or TF as regulators (Figure 1A). In particular, two types of mixed regulatory circuits are identified: (i) a TF that regulates both a given miRNA and its target gene; (ii) a miRNA that regulates both a given TF and its regulated gene (Figure 1B). In order to help the user in the interpretation of the results, novel features are implemented into MAGIA². Among them (i) miRNA–target interactions experimentally validated [as reported in miRecords (28), TarBase (29) and mirTarBase (30) databases] are specifically marked; (ii) Functional enrichment of the gene network component can be performed directly from MAGIA² using DAVID platform.

WEB TOOL IMPLEMENTATION

In Figure 2, MAGIA² architecture comprises three main sections: (i) data upload, (ii) data analysis and methods setup and (iii) results visualization, browsing and linking to external knowledgebase and tools.

Data upload

In the first section, the user selects the organism, defines the gene or transcript ID used (EntrezGene, RefSeq, ENSEMBL gene or transcript are allowed) and uploads miRNA and gene (or transcript) expression data. In this way, a private MAGIA environment is created. Input files must be tab-delimited matrices (samples on the columns and, genes/transcripts or miRNAs on the rows).

Gene and miRNA expression profiles should be already normalized and filtered. Specifically, mRNA and miRNA profiles could be generated using different platforms and could be pre-processed with different algorithms as correlation coefficient and mutual information are invariant under a linear transformation.

Platform-specific probes (Affymetrix and Agilent IDs) are not unique, with several probes mapping on the same gene/transcript and being frequently associated with not concordant expression. Besides, over 30% of the Affymetrix probes do not correctly map to the genes they are supposed to belong (31,32). For these reasons, in the expression signal reconstruction phase, we suggest the user (i) to use custom CDF for Affymetrix technology and (ii) to choose the best probe (in term of quality check flag, or using other criteria) for Agilent technology.

The first columns of matrices should contain miRNA and gene/transcripts IDs of supported organisms; the header lines should be set according to the experimental design chosen. MAGIA² considers two different experimental designs: (i) matched case—gene/transcript and miRNA expression data from the same biological samples and (ii) non-matched case—mRNA/transcript and miRNA expression data on different biological samples, belonging to the same set of classes, possibly resulting in different sample sizes. When using matched profiles the names of the columns of miRNA and gene/transcript matrices should correspond exactly; otherwise,

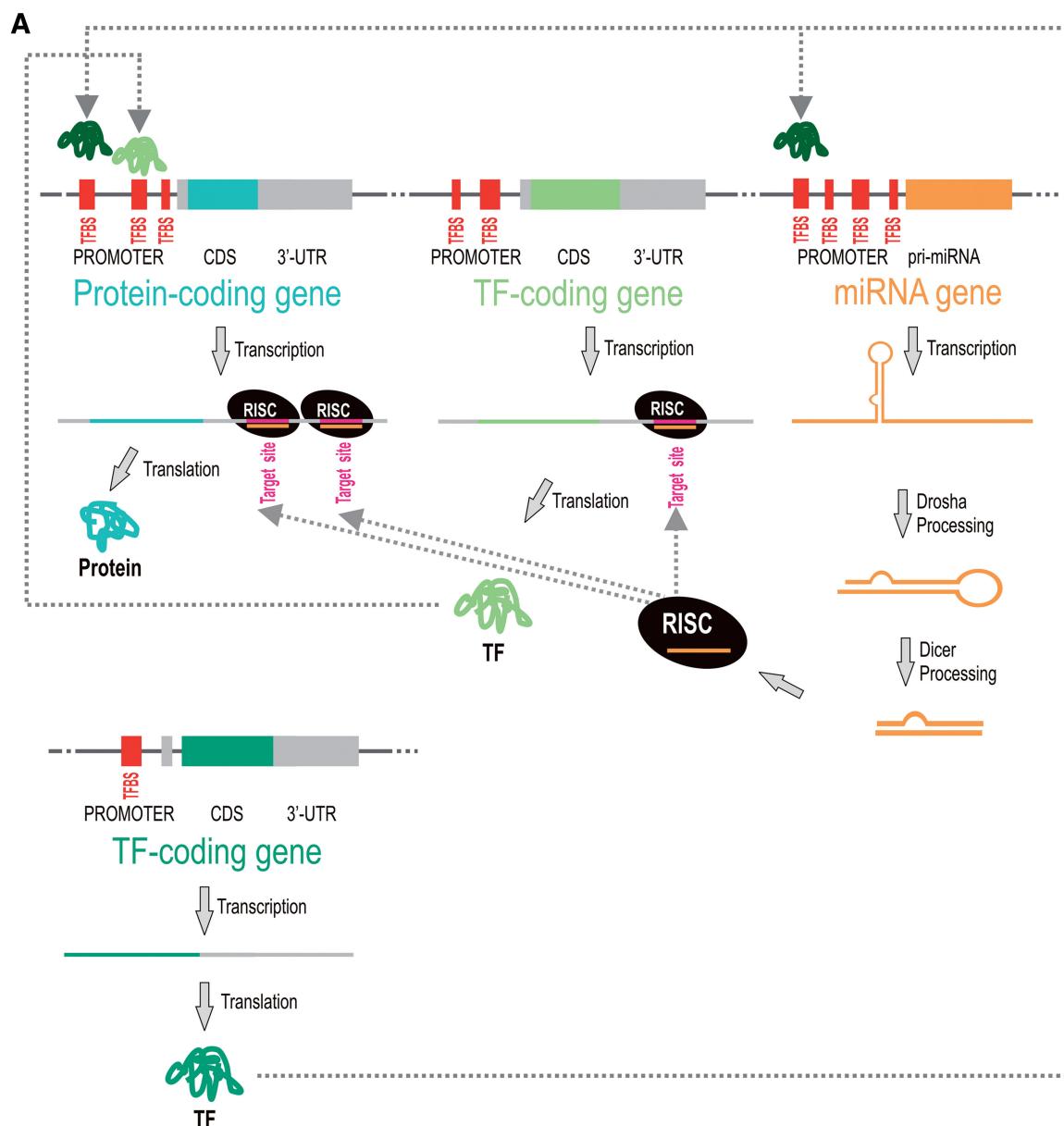
**B**

Figure 1. Interplay of miRNAs and TFs in gene/transcripts expression regulation involves mixed regulatory circuits. Panels A and B represent the molecular biology and the simplified schematization of interconnected pathways of transcriptional and post-transcriptional regulation, respectively: (1) a TF (dark green) interacts both with a protein-coding gene and a miRNA promoter, regulating their transcription, whereas the regulated miRNA targets the protein-coding mRNA exerting post-transcriptional regulation; (2) a miRNA regulates post-transcriptionally both a gene encoding a TF (light green) and a protein-coding gene, that is transcriptional target of the same TF.

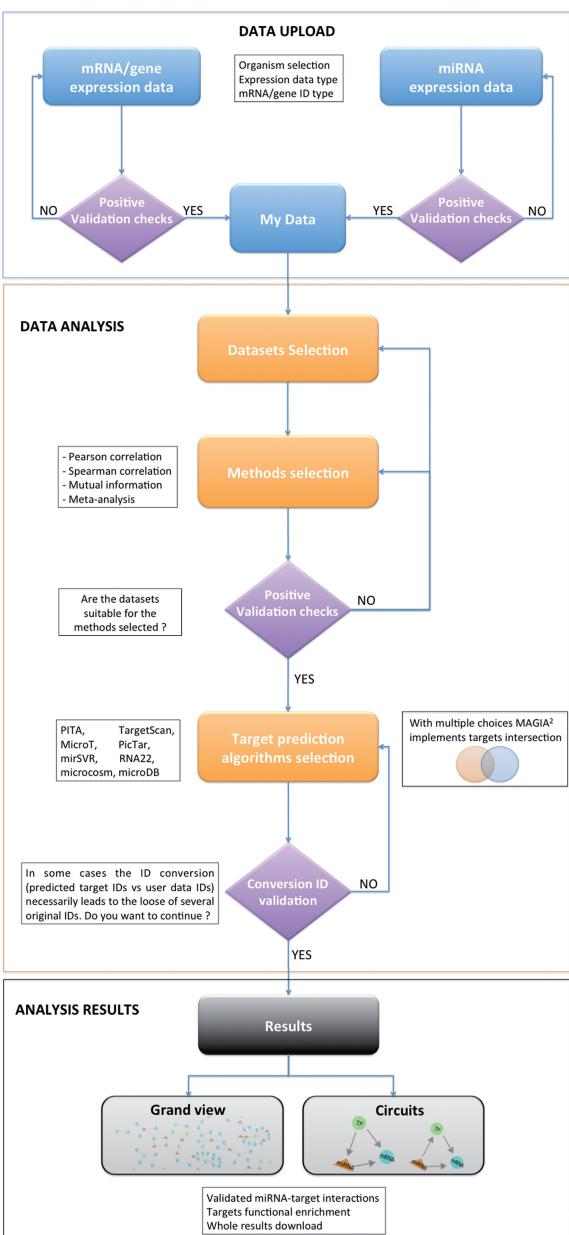


Figure 2. MAGIA² flow chart illustrating the different steps implemented with the back- and front-end in Upload, Analysis and Results sections of MAGIA².

MAGIA² input files require that column labels represent sample classes. When matrices have been successfully uploaded, the user has now created his personal environment.

No registration is required to use MAGIA². To map a user to his personal environment, two standard mechanisms for session management are used: (i) a cookie is set in the user browser; (ii) a session id is appended to each web application url.

The second mechanism allows users to share their personal environment (data and analyses) with collaborators. Moreover, the design of MAGIA² allows two or more users to perform analyses at the same time on

a shared environment without explicit notification or locking.

Analysis setup

After data upload, the user can setup a new analysis using data available in his environment and specify settings in a few simple steps.

Step 1: selection of expression matrices

After creating the 'My data' environment, the user can select two matrices, one for gene/transcripts and one for miRNA expression data. As the presence of miRNAs and genes with poorly variable expression profiles may introduce noise and produce false-positive associations, by default MAGIA² filters profiles according to their expression variability, calculated through the variation coefficient (CV): the 25% less variable profiles (both for miRNAs and for genes/transcripts) are removed. This variability filter can be explicitly skipped using a check box in case the data selected by the user are already filtered. This may apply to a user pre-selected subset of differentially expressed miRNAs that need to be all considered for network reconstruction.

Step 2: selection of the association measure

The selection of the association measures is strictly dependent on the experimental design chosen by the user. For matched design, three methods for expression profiles combination are available (parametric, non-parametric linear correlation and association measure based on information theory), while for un-matched design only a meta-analysis is possible. The Spearman correlation is a non-parametric, rank-based linear correlation measure suitable for non-normally distributed data and/or small sample size (e.g. 3–5). Pearson correlation is a parametric linear correlation measure, suggested for normally distributed data and medium-large sample size (>5). Mutual information is an information measure quantifying the mutual dependence of variables, including non-linear relationships and in general is suitable for large sample size (at least 20 needed).

Meta analysis approach is based on the combination of *P*-values of differential expression (using empirical Bayes test) (33), separately for genes and miRNAs across sample classes, using the inverse chi square distribution to identify oppositely variable miRNA–target pairs (21).

Step 3: selection of the target prediction algorithms

MAGIA² makes use of eight different catalogs of target predictions: Microcosm (34), microrna.org (35), DIANA-microT (36), miRDB (37), PicTar (38), PITA (39), RNA22 (40) and TargetScan (41). Unfortunately not all of them use the same identifiers to refer to genes and transcripts. MAGIA² relieves the user from directly addressing this issue by converting identifiers at the start of each analysis. In case of a conversion requiring multiple steps (e.g. if the user upload matrices with Ensemble gene Ids, their conversion to Refseq would require two steps: Ensemble gene ID—Ensemble transcript ID and Ensemble transcript ID—RefSeq), MAGIA² warns the user. ID conversion, from one transcript ID to another,

in fact usually leads to the loss of a noticeable amount of identifiers and potentially incorrect assignments because there is not a one to one relationship between transcripts identifiers of major sequence databases. Predictions may be further filtered using their scores (MAGIA² tutorial provides information to guide the user in this selection procedure) and multiple targets database may be combined, by taking their intersection.

Integrated analysis: interaction strength quantification and regulatory circuits identification human and mouse data

Using human and mouse data, MAGIA² combines expression profiles analysis with three sets of regulatory interaction predictions, i.e. information about TFs and miRNAs that may regulate a gene/transcript expression at transcriptional and post transcriptional level, respectively, and about transcriptional regulation of TF on miRNAs.

In this way, MAGIA² identifies two types of mixed regulatory circuits: (i) a TF regulating both a given miRNA and its target gene; and (ii) a miRNA regulating both a given TF and its regulated gene.

More in detail, miRNA–target interactions derive from the combination of Steps 2 and 3 above. Namely, only the subset of predicted miRNA–target interactions associated to informative expression profiles relationship measures (e.g. high correlation) are included in the network.

Regarding TFs, we used experimentally validated TF–miRNA interactions reported in mirGen2.0 (17) and TransmiR (42), whereas TF–gene interactions were obtained from ECRbase database (43) (<http://ecrbase.dcode.org/>) for mouse organism and from the ‘TFBS conserved’ track of the UCSC genome annotation for human (version hg19). Each annotated TF binding site was associated with a RefSeq transcript if it lied within 10Kb of its 5'-UTR and 3 Kb of its 3'-UTR. To be more stringent, we required a prediction z-score of at least 3. This resulted in 304,035 total associations: each transcript is linked on average to 11.7 TFBS. By comparison, TargetScan predicts that on average 12.5 microRNAs interact with each transcript.

Thus, association measures described in Step 2 are estimated not only for miRNA–target interactions predicted in Step 3 but also for TF–target relations (TF–gene and TF–miRNA) retrieved from public databases.

Each circuit type might be the result of several different combinations of induction/repression interactions. Using correlation, significant circuits can include different combinations of positive and negative values, highlighting different types of putative regulatory interactions (e.g. TF that inhibits its target mRNA but that activates miRNA expression, a TF that activates both elements). MAGIA² allows the user to download all the circuits in a flat file that can be easily open with a excel sheet for further processing, filtering and analysis and highlight edges according to the sign of the association measure. With this file, the user can check the correlation associated to each edge and speculate on their functional relations.

Rat and *Drosophila* data

Using rat and drosophila data association measures described in Step 2 are estimated only for those miRNA–target interactions predicted in Step 3. TF-target and TF-miRNA information is unavailable.

It is worth noticing that when MAGIA² is used for analysis of miRNA and transcript expression profiles, each alternative transcript of the same gene is treated as a separate entity. In this case, network and circuits visualization and interaction tables report information on individual transcript (and of the gene to which it belongs). The use of transcripts expression increase network complexity but also adds relevant information on transcript heterogeneity, especially when transcript expression is adequately quantified.

MAGIA² RESULTS

MAGIA² results are reported as interactive images and dynamic tables (Figure 3). According to the analysis setting, MAGIA² shows (i) the top 200 functional interactions (between miRNAs, genes and TFs); (ii) top 20 mixed regulatory circuits and (iii) top 40 interactors of a user-selected network node. For each network, two or more tables report different types of supported relationships. In the general view, miRNA–target and TF–target relations are shown, whereas in the circuits visualization mode, triplets of elements involved in type 1 and 2 circuits are listed separately. Focusing on a specific gene/transcript, TF or miRNAs, all direct interactors are shown in appropriate tables. Experimentally validated miRNA–target interactions according to the miRecords (28), miTarbase (29) and miRtarBase (30) are highlighted in tables. Furthermore, for each element involved in a supported interaction, MAGIA² provides links to external miRNA- and gene-related databases to help the user in results interpretation. Complete results can be browsed and download as tab delimited flat files.

From the general network view, a DAVID enrichment analysis can be automatically started. MAGIA² transmits the whole set of genes and TFs represented in the network directly to the DAVID tool for functional enrichment analyses. The default enrichment score provided by DAVID is estimated on the whole genome; however, the user can upload the list of the platform IDs as background to perform a different analysis. This feature may allow the identification of molecular functions under the control of a specific set of transcriptional and post-transcriptional regulators.

CASE STUDY: THE RECONSTRUCTION OF REGULATORY CIRCUITS IN NCI-60 CELL LINES

As a benchmark case study, we used the miRNA and target expression profiles of the NCI-60, a panel of 60 human cancer cell lines from several distinct tissues (44). The NCI-60 expression data were collected using, respectively, the Affymetrix HG-U133A platform and the Ohio State University Comprehensive Cancer Center miRNA platform, OSUCCC, version 3.0. Both data sets are

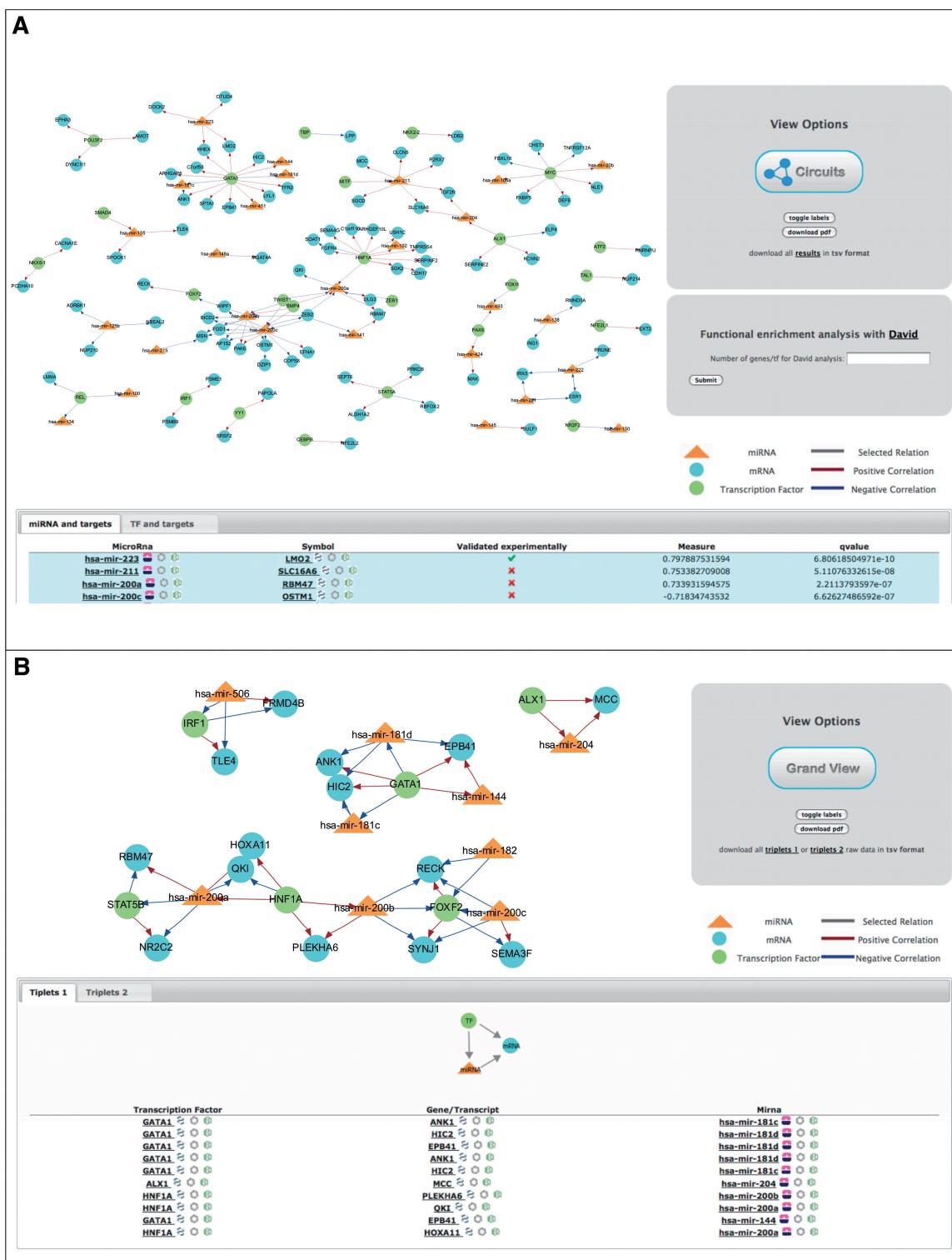


Figure 3. MAGIA² results using NCI-60 matched gene and miRNA expression data. **(A)** Regulatory network reconstructed using the best miRNA-target and TF-target interactions. **(B)** Top 20 mixed regulatory circuits.

available at the ArrayExpress database (mRNA: E-GEOID-5720, miRNA: E-MEXP-1029).

We selected EntrezGene IDs, Pearson correlation measure and the intersection of TargetScan and DIANA-microT target prediction algorithms. We found

a total number of 970 interactions with Pearson correlation $r > 0.4$ in absolute value and with FDR < 0.05 (Table 1). Among these, 13 miRNA–target interactions (two with positive and 11 with negative correlations) are experimentally validated. Given the dense connectivity of

Table 1. Total number of interactions with Pearson correlation $r > 0.4$ in absolute value and with FDR < 0.05 identified by MAGIA² using NCI-60 data set

Interaction type	Pos. corr.	Neg. Corr.	TOT
miRNA–mRNA	216 (44%)	281 (56%)	497
TF–miRNA	14 (46%)	16 (54%)	30
TF–gene	120 (27%)	325 (73%)	444
TOT	350 (36%)	622 (64%)	970

the entire network, MAGIA² shows the sub-network derived from those interactions with $r > 0.4$ or $r < -0.4$, for a maximum of 200 interactions (50 miRNA–gene, 50 miRNA–TF, 50 TF–miRNA and 50 TF–gene). Figure 3 shows the resulting network for this case study (50 miRNA–gene, 1 miRNA–TF, 50 TF–gene and 31 TF–miRNA). Whole results can be downloaded as a tab delimited file with specific attributes that can be directly imported into Cytoscape (45).

miRNA–target interactions include a total number of 17 different miRNAs and of 39 different genes while TF–target interactions include a total number of 27 different TFs and 72 different TF targets. Among the best interactions, we found three miRNA–target relationships experimentally validated as reported in TarBase and miRecords: miR-223 with LMO2 (46) positively correlated with $r = 0.79$, miR-221 with ESR1 and miR-222 with ESR1 (47) inversely correlated with respectively $r = -0.60$ and $r = -0.59$. Interestingly, the interaction between miR-223 with LMO2 is the interaction with the highest score.

Among the top TF–target interactions, we found GATA1 and STAT5A factors. GATA factors are unique TFs with conserved DNA-binding domains. They serve diverse roles in embryogenesis, cell differentiation, regulation of tissue-specific genes and carcinogenesis. Multiple studies have analyzed the role of GATA factors in gastrointestinal malignancy, such as those of the stomach, pancreas and colon, and premalignant lesions such as Barrett's esophagus (48). TF STAT5A, on the other hand, has been shown to play a critical role in prostate and breast cancer growth and differentiation (49).

Pathways enrichment analysis, conducted on target and TF genes, leads to relevant and interesting results: TGF-beta signaling pathway ($P = 0.047$, PANTHER database) (50) and ErbB signaling pathway ($P = 0.032$, KEGG database) (51) are identified as two of the most enriched pathways. Perturbations of transforming growth factor- β (TGF β) signaling are central to tumorigenesis and tumor progression through their effects on cellular process, including cell proliferation and cell invasion (52) while excessive ErbB signaling is associated with the development of a wide variety of types of solid tumor. ErbB-1 and ErbB-2 are found in many human cancers and their excessive signaling may be critical factors in the development and malignancy of these tumors (53).

The most interestingly new feature of MAGIA² is the possibility to identify putative mixed regulatory circuits.

In this analysis (Figure 3B), MAGIA² highlights 20 circuits possibly relevant in tumor samples.

Circuits of type 2, where miRNA acts as master regulator of a TF and its target involve principally FOXF2 and miR-200, miR-506 and miR-204 families. Changes in miR-200 family levels have been associated with enhanced tumorigenesis and resistance to several chemotherapy drugs; in addition, they significantly correlate with decreased survival (54–56).

Among circuits of type 1, where a TF regulates either the miR and its target, we observed an interesting network composed of four circuits (triangles) involving GATA1, as master regulator, miR-144 and miR-181 family, and specific common target genes (EPB41, ANK1 and HIC1). Taken individually, different elements in this net are known to be involved in cancer development and progression. For instance, recent studies showed that miR-181a and miR-181b function as tumor suppressors, which trigger growth inhibition, induce apoptosis and inhibit invasion in glioma cells (57). EPB41 is known as tumor suppressor gene in meningioma pathogenesis (58). The master regulator of hematopoiesis, GATA-1 binds to DNA sites with the consensus sequence [AT]GATA[AG] within regulatory regions of globin genes and of other genes expressed in erythroid cells. It plays also important roles in hematologic malignancies (59) and it seems to directly activate transcription of genes encoding the essential autophagy component (60).

Further, and most interestingly, combined involvements in cancer (as indirect or direct interactions) of miRNAs and specific genes and TFs of this net have been reported.

Specifically, a study of prognostic significance of CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features, and gene and microRNA expression signatures associated with CEBPA mutations (61) showed that up-regulation of specific genes, including GATA1 (ZFPML1, EPOR, and GFI1B) and miRNAs (i.e. the miR-181 family) are involved in erythroid differentiation and down-regulation of homeobox genes. Similarly Whitman *et al.* (62) studied FLT3-internal tandem duplications as adverse prognostic marker in primary cytogenetically normal acute myeloid leukemia and reported an FLT3-ITD-associated expression signature with over-expression of different genes, including ANK1 [FLT3, homeobox genes (MEIS1, PBX3, HOXB3) and immunotherapeutic targets (WT1, CD33) and underexpression of leukemia-associated (MLLT3, TAL1) and erythropoiesis-associated (GATA3, EPOR, ANK1, HEMGN) genes] and underexpression of miR-144.

Focusing more on causal relationships, it was demonstrated that the *miR 144/451* locus, essential for erythropoiesis, is a direct transcriptional target of GATA1 (63). Chromatin immunoprecipitation showed a strong signal for GATA-1 occupancy at the erythroid *cis*-regulatory module (enhancer) located 2.8 upstream of the miR 144/451 transcriptional start site, containing two conserved GATA binding motifs. During erythroid lineage differentiation, GATA-2 binding, characteristic of early precursors, is gradually replaced by GATA-1 that induces gene activation in later stages of maturation.

In summary, we can affirm first that key cancer genes, TFs and miRNAs are represented in top circuits blindly identified by MAGIA² mixed circuits reconstruction based on cancer cell lines expression data. Furthermore, both combined involvement in cancer signatures and/or the existence of direct regulatory interactions relevant for cell differentiation regarding different pairs among the set of TFs, miRNAs and target genes found in MAGIA² mixed circuits clearly corroborates the biological significance of mixed regulatory circuits identified by MAGIA².

ACKNOWLEDGEMENTS

We acknowledge the CINECA Award N. HP10BDJ9X8, 2010 for the availability of high performance computing resources and support.

FUNDING

Funding for open access charge: University of Padova [CPDA119031 to C.R.] to S.B. and to A.C.; Associazione Italiana per la Ricerca sul Cancro (AIRC, Milano) ‘Special Program Molecular Clinical Oncology 5x1000’ to AGIMM (AIRC-Gruppo Italiano Malattie Mieloproliferative).

Conflict of interest statement. None declared.

REFERENCES

- Miller,B.H. and Wahlestedt,C. (2010) MicroRNA dysregulation in psychiatric disease. *Brain Res.*, **1338**, 89–99.
- Maes,O.C., Chertkow,H.M., Wang,E. and Schipper,H.M. (2009) MicroRNA: implications for Alzheimer disease and other human CNS disorders. *Curr. Genomics*, **10**, 154–168.
- Ha,T.-Y. (2011) MicroRNAs in human diseases: from cancer to cardiovascular disease. *Immune Netw.*, **11**, 135–154.
- Eskela-Kerscher,A. and Slack,F.J. (2006) Oncomirs: microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
- O'Connell,R.M., Rao,D.S., Chaudhuri,A.A. and Baltimore,D. (2010) Physiological and pathological roles for microRNAs in the immune system. *Nat. Rev. Immunol.*, **10**, 111–122.
- Rajewsky,N. (2006) microRNA target predictions in animals. *Nat. Genet.*, **38(Suppl)**, S8–S13.
- Flynt,A.S. and Lai,E.C. (2008) Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat. Rev. Genet.*, **9**, 831–842.
- Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Reczko,M. and Hatzigeorgiou,A.G. (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, **25**, 3049–3055.
- Ma,L., Huang,Y., Zhu,W., Zhou,S., Zhou,J., Zeng,F., Liu,X., Zhang,Y. and Yu,J. (2011) An integrated analysis of miRNA and mRNA expressions in non-small cell lung cancers. *PLoS One*, **6**, e26502.
- Nunez-Iglesias,J., Liu,C.-C., Morgan,T.E., Finch,C.E. and Zhou,X.J. (2010) Joint genome-wide profiling of miRNA and mRNA expression in Alzheimer's disease cortex reveals altered miRNA regulation. *PLoS One*, **5**, e8898.
- Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
- Lai,E.C. (2005) miRNAs: whys and wherefores of miRNA-mediated regulation. *Curr. Biol.*, **15**, R458–R460.
- Dorsett,Y., McBride,K.M., Jankovic,M., Gazumyan,A., Thai,T.-H., Robbiani,D.F., Di Virgilio,M., Reina San-Martin,B., Heidkamp,G., Schwickert,T.A. et al. (2008) MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity*, **28**, 630–638.
- Cui,Q., Yu,Z., Pan,Y., Purisima,E.O. and Wang,E. (2007) MicroRNAs preferentially target the genes with high transcriptional regulation complexity. *Biochem. Biophys. Res. Commun.*, **352**, 733–738.
- Inui,M., Martello,G. and Piccolo,S. (2010) MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell Biol.*, **11**, 252–263.
- Wang,J., Lu,M., Qiu,C. and Cui,Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **38**, D119–D122.
- Friard,O., Re,A., Taverna,D., De Bortoli,M. and Corà,D. (2010) CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*, **11**, 435.
- El Baroudi,M., Corà,D., Bosia,C., Osella,M. and Caselle,M. (2011) A curated database of miRNA mediated feed-forward loops involving MYC as master regulator. *PLoS One*, **6**, e14742.
- Le Béchec,A., Portales-Casamar,E., Vetter,G., Moes,M., Zindy,P.-J., Saumet,A., Arenillas,D., Theillet,C., Wasserman,W.W., Lecellier,C.-H. et al. (2011) MIR@NT@N: a framework integrating transcription factors, microRNAs and their targets to identify sub-network motifs in a meta-regulation network model. *BMC Bioinformatics*, **12**, 67.
- Alexiou,P., Vergoulis,T., Gleditsch,M., Prekas,G., Dalamagas,T., Megraw,M., Grossel,I., Sellis,T. and Hatzigeorgiou,A.G. (2010) miRGen 2.0: a database of microRNA genomic information and regulation. *Nucleic Acids Res.*, **38**, D137–D141.
- Sales,G., Coppe,A., Bisognin,A., Biasiolo,M., Bortoluzzi,S. and Romualdi,C. (2010) MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res.*, **38**, W352–W359.
- Nam,S., Li,M., Choi,K., Balch,C., Kim,S. and Nephew,K.P. (2009) MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. *Nucleic Acids Res.*, **37**, W356–W362.
- Liang,Z., Zhou,H., He,Z., Zheng,H. and Wu,J. (2011) mirAct: a web tool for evaluating microRNA activity based on gene expression data. *Nucleic Acids Res.*, **39**, W139–W144.
- Huang,G.T., Athanassiou,C. and Benos,P.V. (2011) mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.*, **39**, W416–W423.
- Friedman,Y., Naamati,G. and Linial,M. (2010) MiRrror: a combinatorial analysis web tool for ensembles of microRNAs and their targets. *Bioinformatics*, **26**, 1920–1921.
- Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Simmoss,V.A., Zhang,L. and Hatzigeorgiou,A.G. (2010) The DIANA-mirExTra web server: from gene expression data to microRNA function. *PLoS One*, **5**, e9171.
- Nam,S., Kim,B., Shin,S. and Lee,S. (2007) miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.
- Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
- Vergoulis,T., Vlachos,I.S., Alexiou,P., Georgakilas,G., Maragkakis,M., Reczko,M., Geranelos,S., Koziris,N., Dalamagas,T. and Hatzigeorgiou,A.G. (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.*, **40**, D222–D229.
- Hsu,S.-D., Lin,F.-M., Wu,W.-Y., Liang,C., Huang,W.-C., Chan,W.-L., Tsai,W.-T., Chen,G.-Z., Lee,C.-J., Chiu,C.-M. et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **39**, D163–D169.
- Ferrari,F., Bortoluzzi,S., Coppe,A., Sirota,A., Safran,M., Shmoish,M., Ferrari,S., Lancet,D., Danieli,G.A. and Bicciato,S. (2007) Novel definition files for human GeneChips based on GeneAnnot. *BMC Bioinformatics*, **8**, 446.
- Dai,M., Wang,P., Boyd,A.D., Kostov,G., Athey,B., Jones,E.G., Bunney,W.E., Myers,R.M., Speed,T.P., Akil,H. et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.*, **33**, e175.

33. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article 3.
34. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
35. Betel,D., Wilson,M., Gabow,A., Marks,D.S. and Sander,C. (2007) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
36. Maragkakis,M., Alexiou,P., Papadopoulos,G.L., Reczko,M., Dalamagas,T., Giannopoulos,G., Goumas,G., Koukis,E., Kourtis,K., Simossis,V.A. et al. (2009) Accurate microRNA target prediction correlates with protein repression levels. *BMC Bioinformatics*, **10**, 295.
37. Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
38. Krek,A., Grun,D., Poy,M.N., Wolf,R., Rosenberg,L., Epstein,E.J., MacMenamin,P., da Piedade,I., Gunsalus,K.C., Stoffel,M. et al. (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
39. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
40. Miranda,K.C., Huynh,T., Tay,Y., Ang,Y.-S., Tam,W.-L., Thomson,A.M., Lim,B. and Rigoutsos,I. (2006) A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.
41. Grimson,A., Farh,K.K.-H., Johnston,W.K., Garrett-Engele,P., Lim,L.P. and Bartel,D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell*, **27**, 91–105.
42. Wang,J., Lu,M., Qiu,C. and Cui,Q. (2010) TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **38**, D119–D122.
43. Loots,G.G. and Ovcharenko,I. (2007) ECRbase: database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics*, **23**, 122–124.
44. Weinstein,J.N. (2006) Spotlight on molecular profiling: “Integromic” analysis of the NCI-60 cancer cell lines. *Mol. Cancer Ther.*, **5**, 2601–2605.
45. Smoot,M.E., Ono,K., Ruscheinski,J., Wang,P.L. and Ideker,T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
46. Yuan,J.-Y., Wang,F., Yu,J., Yang,G.-H., Liu,X.-L. and Zhang,J.-W. (2009) MicroRNA-223 reversibly regulates erythroid and megakaryocytic differentiation of K562 cells. *J. Cell. Mol. Med.*, **13**, 4551–4559.
47. Zhao,J.-J., Lin,J., Yang,H., Kong,W., He,L., Ma,X., Coppola,D. and Cheng,J.Q. (2008) MicroRNA-221/222 negatively regulates estrogen receptor alpha and is associated with tamoxifen resistance in breast cancer. *J. Biol. Chem.*, **283**, 31079–31086.
48. Ayanbule,F., Belaguli,N.S. and Berger,D.H. (2011) GATA factors in gastrointestinal malignancy. *World J. Surg.*, **35**, 1757–1765.
49. Koptyra,M., Gupta,S., Talati,P. and Nevalainen,M.T. (2011) Signal transducer and activator of transcription 5a/b: biomarker and therapeutic target in prostate and breast cancer. *Int. J. Biochem. Cell Biol.*, **43**, 1417–1421.
50. Thomas,P.D., Kejariwal,A., Campbell,M.J., Mi,H., Diemer,K., Guo,N., Ladunga,I., Ulitsky-Lazareva,B., Muruganujan,A., Rabkin,S. et al. (2003) PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res.*, **31**, 334–341.
51. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
52. Blobe,G.C., Schiermann,W.P. and Lodish,H.F. (2000) Role of transforming growth factor beta in human disease. *N. Engl. J. Med.*, **342**, 1350–1358.
53. Normanno,N., De Luca,A., Bianco,C., Strizzi,L., Mancino,M., Maiello,M.R., Carotenuto,A., De Feo,G., Caponigro,F. and Salomon,D.S. (2006) Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*, **366**, 2–16.
54. Marchini,S., Cavalieri,D., Frusci,R., Calura,E., Garavaglia,D., Nerini,I.F., Mangioni,C., Cattoretti,G., Clivio,L., Beltrame,L. et al. (2011) Association between miR-200c and the survival of patients with stage I epithelial ovarian cancer: a retrospective study of two independent tumour tissue collections. *Lancet Oncol.*, **12**, 273–285.
55. Pogribny,I.P., Filkowski,J.N., Tryndyak,V.P., Golubov,A., Shpyleva,S.I. and Kovalchuk,O. (2010) Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *Int. J. Cancer*, **127**, 1785–1794.
56. Iorio,M.V., Visone,R., Di Leva,G., Donati,V., Petrocca,F., Casalini,P., Taccioli,C., Volinia,S., Liu,C.-G., Alder,H. et al. (2007) MicroRNA signatures in human ovarian cancer. *Cancer Res.*, **67**, 8699–8707.
57. Shi,L., Cheng,Z., Zhang,J., Li,R., Zhao,P., Fu,Z. and You,Y. (2008) hsa-mir-181a and hsa-mir-181b function as tumor suppressors in human glioma cells. *Brain Res.*, **1236**, 185–193.
58. Piaskowski,S., Rieske,P., Szybka,M., Wozniak,K., Bednarek,A., Pluciennik,E., Jaskolski,D., Sikorska,B. and Liberski,P.P. (2005) GADD45A and EPB41 as tumor suppressor genes in meningioma pathogenesis. *Cancer Genet. Cytogenet.*, **162**, 63–67.
59. Bresnick,E.H., Katsumura,K.R., Lee,H.Y., Johnson,K.D. and Perkins,A.S. (2012) Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res.*, April 5 (doi:10.1093/nar/gks281; epub ahead of print).
60. Kang,Y.A., Sanalkumar,R., O'Geen,H., Linnemann,A.K., Chang,C.J., Bouhassira,E.E., Farnham,P.J., Keles,S. and Bresnick,E.H. (2012) Autophagy driven by a master regulator of hematopoiesis. *Mol. Cell Biol.*, **32**, 226–239.
61. Marcucci,G., Maharry,K., Radmacher,M.D., Mrózek,K., Vukosavljevic,T., Paschka,P., Whitman,S.P., Langer,C., Baldus,C.D., Liu,C.G. et al. (2008) Prognostic significance of, and gene and microRNA expression signatures associated with, CEBPA mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a cancer and leukemia group B study. *J. Clin. Oncol.*, **26**, 5078–5087.
62. Whitman,S.P., Maharry,K., Radmacher,M.D., Becker,H., Mrózek,K., Margeson,D., Holland,K.B., Wu,Y.Z., Schwind,S., Metzeler,K.H. et al. (2010) FLT3 internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *Blood*, **116**, 3622–3626.
63. Dore,L.C., Amigo,J.D., Dos Santos,C.O., Zhang,Z., Gai,X., Tobias,J.W., Yu,D., Klein,A.M., Dorman,C., Wu,W. et al. (2008) A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc. Natl Acad. Sci. USA*, **105**, 3333–3338.